Express analysis of protein amino acid sequences

Primary structure of *Penicillium chrysogenum* 152A guanyl-specific ribonuclease

S.V. Shlyapnikov, S.I. Bezborodova*, V.A. Kulikov and G.I. Yakovlev

Institute of Molecular Biology, the USSR Academy of Sciences and **Institute of Genetics and Selection of Industrial Microorganisms, Vavilova 32, Moscow 117312, USSR*

Received 11 November 1985

The complete amino acid sequence of *Penicillium chrysogenum* 152A guanyl-specific RNase has been established using automated Edman degradation of two non-fractionated peptide mixtures produced by tryptic and staphylococcal protease digests of the protein. The RNase contains 102 amino acid residues: His_2, Arg_3, Asp_7, Asn_8, Thr_5, Ser_{11}, Glu_4, Gln_2, Pro_4, Gly_{11}, Ala_{13}, Cys_4, Val_8, Ile_3, Leu_3, Tyr_9, Phe_5 (M_r 10 747).

RNase (Penicillium chrysogenum) *Primary structure*

1. INTRODUCTION

Amino acid sequencing of proteins is a rather complicated problem of which the solution depends greatly on the skill of the worker and on the choice of an optimal research strategy. Progress made recently, first of all, in automated Edman polypeptide degradation in combination with highly sensitive techniques used for separation of amino acid derivatives calls for elaboration of more effective approaches in studies of protein primary structures.

This paper describes the express method of amino acid sequencing of *Penicillium chrysogenum* 152A guanyl-specific RNase (RNase Pch_1), which may be useful for studying the primary structures of other proteins.

2. MATERIALS AND METHODS

RNase Pch_1 was isolated from the fungus *P. chrysogenum* 152A [1,2], reduced and carboxymethylated as in [3].

The RNase was digested with staphylococcal protease V8 and trypsin in 0.1 M ethylmorpholine acetate buffer, pH 8.1, at 37°C for 8 or 4 h and at a final enzyme/substrate ratio of 1:30 or 1:50, respectively. In the latter case, an S-Cm-RNase preparation preliminarily subjected to 8 complete cycles of automated Edman degradation was digested in a sequencer reaction cell.

The kinetics of RNase hydrolysis with carboxypeptidase was followed up by analysing aliquots of the reaction mixture (0.05 M ethylmorpholine acetate, pH 5.5, in the presence of 1% SDS at a 1:100 enzyme/substrate ratio) with a Biotronik LC 7000 amino acid analyser.

Automated Edman degradation was performed with a Beckman 890C sequencer according to the 0.1 M Quadrol program in the presence of polybrene [4,5]. Reagents and solvents were from Beckman.

PTH-amino acid derivatives were identified by gas-liquid [6] and thin-layer chromatography [7] as well as by amino acid analysis after back-hydrolysis of the products with 5 N HI at 150°C for 24 h. Edman degradation derivatives were converted in 1 N HCl [8]. PTH-Nle was used as an internal standard when the yield of sequencer products was analysed quantitatively.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

3. RESULTS AND DISCUSSION

The principle of determining the primary structure of RNase Pch_1 is based on using the technique of automated sequencing two non-fractionated peptide hydrolysates of the protein, namely, the products of RNase proteolysis by staphylococcal protease and trypsin. Such a strategy adapts a

Table 1

Automated Edman degradation of S-Cm-RNase and des-(1–8)-S-Cm-RNase Pch_1 peptide hydrolysates

Degradation cycle	Yield of amino acids (nmol)										
	Tryptic hydrolysate of des-(1–8)-RNase				Hydrolysate of RNase produced by staphylococcal protease						
1	2				3						
1	Val 56G	Asn 18A	Ser 24A		Ala 62G	Leu 19G	Tyr 18G	Gly 20G	Phe 21G		
2	Cys(Cm) 20A	Tyr 23G	Gly 20G	Val 29G	Cys(Cm) G,T	Gly 15G	Tyr 6G	Arg A,T	Phe 31G	Pro 20G	Ala 8G
3	Tyr 20G	Glu 27A	Ala 25G	Phe 24G	Ala 41G	Tyr 10G	Ser G,T	Asn G,T	Asp 40A	Ile 21G	Gly 5G
4	Thr 14A	Gly 19G	Val 25G	Asn 13A	Ala 47G	Asp 8A	Tyr 18G	Phe 25G	Leu 17G	Val 6G	
5	Ser 18A	Phe 17G	Tyr 19G	Gly 18A	Thr 15A	Leu 13A	Asn 3A	Glu 10A	Pro 19G	Arg A	Ile 7A
6	Ser 40A	Asp 35A	Asn G,T		Cys(Cm) G,T	Tyr 8G	Asp 5A	Val 26G	Ser G,T	Thr 2A	
7	Ala 22G	Phe 16G	Gly 14G	Asp 18A	Gly 33G	Ser 20A	Asp 6A	His A			
8	Ile 18G	Pro 19G	Asn 15A	Gln 12A	Ser G,T	Ala 32G	Val 7G	Gly 19A	Thr 3A		
9	Ser 40A	Val 22G	Leu 16G		Val 52G	Asn 7A	Ser 4A	Thr 10A	Gly 4G		
10	Ala 33G	Ser G,T	Pro 14G		Cys(Cm) G,T	Asp 15A	Asn G,T	Tyr 25G	Ala 5G		
11	Ala 16G	Gly 10G			Tyr 36G	Asp 11A	Ser 11A				
12	Gln 10A	Thr 10A	Ala 18G	Val 13G	Thr 9A	Val 10G	Pro 4G	Glu 5A	Gly 19A		
13	Glu 15A	Tyr 18G	Asp 15A	Ile 12G	Ser 15A	His A	Asn 14A	Phe 5G			
14	Ala 19G	Tyr 14G	Arg A	Thr 10A	Ser 24A	Asn 16A	Glu 2A	Pro 4G			
15	Gly 7G	Glu 12A	His A		Ala 25G	Tyr 6G	Pro 10G	Phe 3G	Ile 6G		
16	Tyr 13G	Phe 12G	Thr 8A		Ile 22A	Pro 8G	Gly 11A	Val 2G	Leu 4A		

Table 1 (continued)

Degradation cycle	Yield of amino acids (nmol)								
	Tryptic hydrolysate of des-(1–8)-RNase			Hydrolysate of RNase produced by staphylococcal protease					
1	2			3					
17	Asp 10A	Pro 9G	Gly 8G	Ser G,T	His A	Ala 16G	Arg A		
18	Leu 12A	Ile 8A	Ala 9G	Ala 18G	Glu 3A	Asp 15A	Ser G,T	Cys(Cm) G,T	
19	Tyr 7G	Leu 10G	Ser 7A	Ala 14G	Tyr 3A	Arg A	Gly 3G		
20	Ser 11A	Arg A	Gly 8G	Gln 3A	Arg A	Val 14G	Ala 3G		
21	Ala 12G	Asn 8A		Glu 2A	Asn 2A	Val 19G			
22	Asn 13A			Phe 11G	Tyr 3A				
23	Asp 9A	Phe 7G		Asn 10A					
24	Asp 10A	Val 6G		Gly 8A					
25	Val 8G	Ala 7G		Asn 7A					
26	Cys(Cm) G,T	Ser G,T		Asp 8A					
27	Asn 2A			Gln 5A					
28	Tyr 6G			Leu 6G					
29	Pro 4G			Ala 4G					
30	His A			Gly 3G					
31	Glu 2A			Val 4G					
32	Tyr 3A			Ile 2A					
33	Arg A			– –					

The yield of amino acids (without background) was determined by gas-liquid (G) and thin-layer (T) chromatography and amino acid analysis (A). The level of background for each amino acid was summarized from general 'chemical noise' (which was considered equal to 4 and 2 nmol for the identification methods A and G, respectively) and the value 'residual amino acid overlapping', calculated as 10% of the amount amino acid at the previous step of the degradation

general theoretical methodology of the protein primary structure express analysis, which was suggested by Gray [9], and is a logical sequel of our earlier study with homologous *P. brevicompactum* RNase whose amino acid sequence was investigated without separation and purification of individual peptides [5,10]. The obvious homology of the two fungal guanyl-specific RNases is confirmed by the similarity of their amino acid composition [1,5] as well as of the active-site structure and the mechanism of functioning [11,12]. At the same time, the present work supposes the use of a highly formalized approach in which preliminary information about a protein structure is minimized and takes into consideration in essence the approximate efficiency of homologous RNases proteolysis with similar proteases. The main special feature of RNase Pch_1 enzyme digestions as compared to those in the case of *P. brevicompactum* RNase [5] was carrying out tryptic proteolysis on the preparation of S-Cm-RNase preliminarily subjected to 8 complete steps of automated sequencing – des-(1–8)-RNase Pch_1. Such a modification made it possible to restrict the selectivity of trypsin cleavage at Arg residues [1,2].

Four polypeptide chains undergo Edman degradation in the tryptic hydrolysate of des-(1–8)-RNase Pch_1 (see table 1). This corresponds to a content of 3 Arg residues in the protein and shows that the yield of peptides in the proteolysis is nearly quantitative.

The composition of the peptide hydrolysate generated by RNase Pch_1 digestion with staphylococcal protease is more complex. Specific proteolysis of the protein with this enzyme at 4 Glu residues as well as additional fragmentation of the peptide chain at the N-terminal bonds of 3 Leu residues detected at the N-termini of peptides in the hydrolysate and absent at steps 4, 14, 16 and 32 in Edman degradation of des-(1–8)-RNase tryptic peptides suggest that the original mixture contains up to 8 components. The noticeable dispersion in the yield of amino acid derivatives and the identification of 5 Glu residues, when the hydrolysate produced by RNase digestion with staphylococcal protease was sequenced, stem from incomplete protein proteolysis (table 1).

The algorithm of RNas Pch_1 primary structure determination included the combined analysis of products generated in Edman degradation of RNase and des-(1–8)-RNase peptide hydrolysates, the protein N-terminal amino acid sequence Ala-Cys(Cm)-Ala-Ala-Thr-Cys(Cm)-Gly-Ser established upon 8 steps of intact S-Cm-RNase automated degradation and the C-terminal sequence Phe-Val-Ala-Cys(Cm)-Asp derived from considering the kinetics of the protein hydrolysis with carboxypeptidase Y. Here, products generated at cycle 1 in Edman degradation of the hydrolysate obtained upon RNase digestion with staphylococcal protease were compared with those produced at steps 4, 14, 16 and 32 in sequencing des-(1–8)-RNase tryptic peptides directly adjacent to Glu-containing cycles as well as with amino acids identified at steps 9, 18 and 19 of the respective hydrolysate containing Leu; products yielded at cycle 2 were compared with those of cycles 5, 15, 33, 10, 19 and 20, etc. When Arg residues were detected in the studied steps of Edman degradation of the RNase hydrolysate produced by staphylococcal protease, the data of sequencing the tryptic hydrolysate of des-(1–8)-RNase beginning from the first cycle of degradation were additionally included in the analysis of the next step.

The absence of coincidences in the 'reading frames' for amino acid sequences of peptides in the RNase Pch_1 hydrolysate produced by staphylococcal protease, which are specified by localisation of residues Glu and Leu in the consecutive cycles of sequencing the tryptic hydrolysate of des-(1–8)-protein, allows one to analyse merely two peptide mixtures. At the same time, combined analysis of amino acid sequences of all the peptides by RNase digestion with staphylococcal protease together with quantitative determination of amino acid derivatives detected at the corresponding steps of Edman degradation of the tryptic peptides suffices for unambiguous interpretation of local 'degeneration' in the identification of amino acids in individual positions of certain peptide structures, for the mutual arrangement of peptides and therefore for unambiguous elucidation of the complete primary structure of a protein (fig.1).

Among the 6 peptide bonds cleaved in the molecule of RNase Pch_1 with staphylococcal protease (see fig.1), only Glu-Ala (peptides 1/2) and Glu-Gly (peptides 2+4/5) bonds are hydrolysed with a virtually quantitative yield. Minor proteolysis of Asp-Leu (peptide 3) and Gln-Leu (peptide 7) bonds, on the one hand, stems from non-

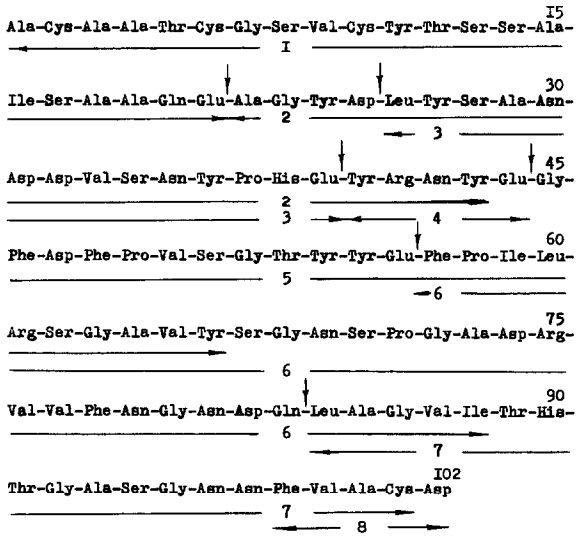

Fig.1. RNase Pch_1 amino acid sequence. The sites of protein cleavage with staphylococcal protease are indicated by vertical arrows. (1–7) Established structures of peptides in the hydrolysate of RNase produced by staphylococcal protease; (8) data from studying the kinetics of RNase hydrolysis with carboxypeptidase Y.

absolute specificity of the protease towards Glu residues and, on the other, from partial desamidation of Gln^{83} in the protein. The low degree of hydrolysis (60%) of Glu-Tyr and Glu-Phe bonds (peptides 4 and 6) may be caused by the fact that the cleavage is inhibited by the residues of hydrophobic amino acids and closely located Pro [13,14].

An appraisal of our technique for analysing the amino acid sequence of RNase Pch_1 shows the advantage of such a strategy which allows one to establish the entire primary structure of the protein within a very short period of time using a minimal quantity of the protein. The solution of the problem has been facilitated considerably because the object under study was comparatively simple. This has made it possible to perform the analysis without computers and, even more significantly, to reach a high efficiency in the automated sequencing of peptide mixtures at a low artifact background.

The method offers the promise of being applicable to a wide range of other proteins whose primary structure is to be analysed. Here, one has to choose a proper strategy for the selective fragmentation of a polypeptide chain using, as a criterion, the overall number of Edman degradation steps necessary for analysing the entire structure. Progress in the application of Edman degradation, its instrumental and technological equipment, process control using highly sensitive amino acid analysis by liquid chromatography and computer processing of results allow one to hope that the proposed technique will be adapted for further studies of protein primary structures.

REFERENCES

[1] Bezborodova, S.I., Markelova, N.Yu. and Guljaeva, V.I. (1975) Biokhimiya 40, 592–597.
[2] Bezborodova, S.I., Beletzkaya, O.P. and Gritchenko, V.M. (1977) Biokhimiya 42, 1556–1566.
[3] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 622–634.
[4] Klapper, D.G., Wilde, C.E. iii and Capra, J.D. (1978) Anal. Biochem. 85, 126–131.
[5] Kulikov, V.A., Shlyapnikov, S.V. and Yakovlev, G.I. (1985) Bioorg. Khim. 11, 334–344.
[6] Pisano, J.J., Bronzert, T.J. and Brewer, H.B. (1972) Anal. Biochem. 45, 43–59.
[7] Alakhov, Yu.B., Motuz, L.P., Stengrevicz, O.A. and Vinokurov, L.M. (1978) Bioorg. Khim. 4, 1301–1313.
[8] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[9] Gray, W.R. (1968) Nature 220, 1300–1304.
[10] Shlyapnikov, S.V., Kulikov, V.A. and Yakovlev, G.I. (1984) FEBS Lett. 177, 246–248.
[11] Yakovlev, G.I., Karpeisky, M.Ya., Bezborodova, S.I., Beletskaya, O.P. and Sakharovsky, V.A. (1980) Eur. J. Biochem. 109, 75–85.
[12] Karpeisky, M.Ya., Yakovlev, G.I., Bott, V., Ezhov, V.A. and Prikhodko, A.G. (1981) Bioorg. Khim. 7, 1335–1347.
[13] Houmard, J. and Drapeau, G.R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[14] Austen, B.M. and Smith, E.L. (1976) Biochem. Biophys. Res. Commun. 72, 411–417.